

Übungen zur Vorlesung Sequenzanalyse

Universität Bielefeld, SS 2020

Prof. Dr. Jens Stoye · Dr. Marília D. V. Braga · Leonie R. Brockmann · Rebecca K. Pfeil

<https://gi.cebitec.uni-bielefeld.de/teaching/2020summer/sa>

Übungsblatt 2 vom 30.4.2020

Abgabe am 7.5.2020 bis 12:00 Uhr (mittags)

Aufgabe 1 (*Edit-Operationen*)

(2 Punkte)

Berechne den String $S = E(NXWZLOKP, \mathcal{CS}_E \mathcal{CDS}_Y \mathcal{CI}_R \mathcal{CD})$. Gib alle Zwischenschritte an. Wie hoch sind die Kosten der Edit-Sequenz?

Aufgabe 2 (*Edit-Sequenz und -Distanz*)

(6 Punkte)

Überführe den String $x = \text{FRUEHLINGSZEIT}$ in den String $y = \text{BLUETENKLEID}$ durch Anwendung der *Edit*-Funktion.

1. Gib eine *Edit*-Sequenz mit minimalen Kosten an, die x in y überführt. Wie viele Edit-Sequenzen mit minimalen Kosten gibt es? Sind die *Edit*-Sequenzen mit minimalen Kosten immer die kürzesten?
2. Gib eine *Edit*-Sequenz mit maximalen Kosten an, die x in y überführt. Wie viele Edit-Sequenzen mit maximalen Kosten gibt es? Sind die *Edit*-Sequenzen mit maximalen Kosten immer die längsten?

Aufgabe 3 (*Berechnung der Edit-Distanz*)

(4 Punkte)

Gegeben seien die Strings $x = \text{BANANA}$ und $y = \text{TANNE}$. Benutze zur Berechnung der Edit-Distanz der beiden Strings eine Edit-Matrix und gib die Distanz an. Benutze außerdem eine weitere Matrix, in der die optimalen Edit-Operationen gespeichert werden, um alle optimalen Edit-Sequenzen bestimmen zu können. Schreibe eine optimale Edit-Sequenz explizit auf.

Aufgabe 4 (*Edit-Distanzen*)

(8 Punkte)

Die Rekurrenz zur Berechnung der Standard-Edit-Distanz mit Einheitskosten lautet für $1 \leq i \leq |x|, 1 \leq j \leq |y|$:

$$D(i, j) = \min \begin{cases} D(i-1, j-1) + \mathbb{1}_{\{x[i] \neq y[j]\}} \\ D(i-1, j) + 1 \\ D(i, j-1) + 1 \end{cases}$$

Die Rekursionsbasis ist gegeben durch:

$$D(0, j) = j \text{ für } 0 \leq j \leq |y| \text{ und } D(i, 0) = i \text{ für } 0 \leq i \leq |x|$$

1. Wie sehen die Rekurrenz und die Basisfälle für die *Hamming-Distanz* aus?
2. Wie sehen die Rekurrenz und die Basisfälle für die *LCS-Distanz* aus?
3. Ist es möglich, die *Edit+Flip-Distanz* rekursiv zu berechnen? Wenn nein, warum nicht? Wenn ja, gib eine Formel und eine Rekursionsbasis an. Welche Schwierigkeiten könnten zudem hierbei auftreten?